Filtering Spam "At Your Leisure:" A MAAWG Briefing/Look Ahead

MAAWG 7th General Meeting
Brussels, Belgium, 27-29 June 2006

Joe St Sauver, Ph.D. (joe@uoregon.edu)
MAAWG Senior Technical Advisor
http://www.uoregon.edu/~joe/maawg7

I'm Delighted to Be Here in The Heart of the EU Today, Truly On The Messaging "Front Lines"…

- Some Americans (including me?) are prone to provincialism, somehow implicitly (and mistakenly) assuming that email in general (or spam in particular) is primarily a "US thing."
- In reality, if you look at the top domains sourcing email traffic worldwide (both good and bad) as measured by Senderbase, you'll see that nearly **50%** of the email traffic coming from "top talker" domains now comes from **EUROPE**.
- By way of example, see the following slide for estimated daily message volume for a typical day (used here courtesy Senderbase). European providers have been bolded in that table.

SENDERBASE TOP ESTIMATED DAILY MAIL VOLUME, LAST 24 HOURS (23 JUN 06)

419.5 million	**tpnet.pl**	**Polish Telecom**
312.3	**rima-tde.net**	**Telephonic de Espana**
219.0	cndata.com	CHINANET backbone network
206.3	comcast.net	Comcast Cable
206.3	**proxad.net**	**Proxad / Free SAS**
203.6	rr.com	Road Runner
203.2	yahoo.com	Yahoo!
198.9	verizon.net	Verizon Internet Services
183.8	**interbusiness.it**	**TELECOM ITALIA S.P.A.**
176.7	hinet.net	CHTD, Chunghwa Telecom Co., Ltd.
171.7	**wanadoo.fr**	**France Telecom IP2000 ADSL BAS**
135.5	telesp.net.br	TELECOMUNICACOES DE SAO PAULO
96.8	veloxzone.com.br	Telemar Norte Leste S.A.
95.7	charter.com	CHARTER COMMUNICATIONS
94.3	**ttnet.net.tr**	**Turk Telekom**
88.2	hotmail.com	MS Hotmail
78.4	**bezeqint.net**	**ADSL-CUSTOMER-CONNECTION**
75.5	**t-dialin.net**	**Deutsche Telekom AG**
72.5	**gaoland.net**	**Dynamic pool**
72.1	brasiltelecom.net.br	Brasil Telecom S/A

1614.3 Million European / 3310.3 Million Total ==> Top Mail Volume Is *48.7% European*

Good and Bad Traffic…

- Let me be clear: much of the mail volume shown in the preceding table is legitimate. Some, equally unquestionably, is spam or other types of unwanted traffic (such as phishing email, email borne malware, etc.) -- this is true for European and non-European providers alike.
- However, since European messaging providers are such a major part of the email community these days, it is critical that all leading messaging providers work together to share best practices to pragmatically and cost effectively manage spam while maintaining the assured deliverability of the mail their customers want.
- For that reason, as well as for many others, I'm happy to be here with you in Brussels today, and I hope we will see a steady influx of new MAAWG members from Europe.

Earlier Talks

- Previously I've talked about how carriers can scalably deal with zombies (http://www.uoregon.edu/~joe/zombies.pdf), and at the last MAAWG I talked about:
  (1) the value of filtering with SURBLs,
  (2) SMTP Auth, Port 587 and encryption,
  (3) email traffic from high density shared hosting providers,
  (4) spam filtering complexity, and
  (5) U.S. federal enforcement of CAN-SPAM
  (see http://www.uoregon.edu/~joe/maawg6/ for that talk).
- There are still many other things we could talk about today, but in order to give the other panelists a shot at some time, for a change I'll limit myself to just ONE topic, filtering spam "at your leisure."

Talk Format and Disclaimer

- The content of this talk has been carefully tailored for a mixed managerial/technical "public" audience.
- That said, these slides are quite detailed. Why?
  -- Time is limited and I'll get "side tracked" and run late if I don't "stick to the script"
  -- I usually cover quite a bit of material fairly quickly
  -- I hate to be misquoted
  -- I like to provide pointers to sources for further information (but hate to make you all frantically scribble URLs)
  -- I know these slides may be viewed after the fact by those who are not here in person today, and also by those for whom English is not their primary language; it may help to think of these slides as "closed captioning" for my talk.
- **Disclaimer:** all opinions expressed in this talk are strictly my own. It would be really foolish to act on anything I suggest without doing your own "due diligence" first.

Filtering Spam "At Your Leisure"

"A Miniscule Window for Scrutiny"

- In the traditional/currently dominant spam filtering paradigm, filtering happens (or doesn't happen) at mail delivery time, ideally while the remote system is still connected, so that unwanted mail can be rejected rather than bounced to a likely-forged apparent sender address.
- Unfortunately, that paradigm has some limitations…
  -- you may receive traffic from previously unknown systems, with the result that you have to make an accept/reject decision with little or no reputation data (yet) available
  -- you may have only limited time to scrutinize the message headers and message body (new messages are coming in all the time, so you can't computationally grind on a message for a protracted period of time)
  -- categorization of a message as spam/ham is effectively irrevocable, and not subject to later reconsideration

Providers End Up "Racing" Spammers…

- Spammers are aggressively working to deliver mail before they (and their message stream) gets known/blocked.
- At the same time, providers are accumulating user complaints as well as in house and third party reputation data, adaptively tweaking filters in an effort to block spam while it is still "in flight" (while simultaneously working hard to avoid accidentally blocking real mail)
- Predictably, with the spammers pushing hard and the providers limited by stringent time constraints, at least some spam ends up getting delivered (and some real email ends up getting accidentally blocked).
- But we're talking about a fundamentally crazy paradigm. **Email is asynchronous**, yet we've been unnecessarily constraining ourselves by treating email as if it were synchronous. It's not!

An Alternative Paradigm….

- While it's convenient to filter mail while the remote system is still connected (or shortly thereafter), we actually have a tremendous amount of time to filter spam more or less "at our leisure," post-delivery, right up until the time the user finally reads that message:

  >=========>==========> [minutes/hours/days…] ======>

connect time	mail transferred and spooled/delivered	user finally reads msg

- ***MESSAGE DELIVERY ATTEMPTS SHOULD GENERALLY *NOT* BE CONSIDERED FINAL UNLESS THE MESSAGE HAS BEEN REJECTED OR DISCARDED, OR THE MESSAGE HAS BEEN OPENED BY THE USER.***

What Might Happen During Our Expanded Decision Window?

- Many message characteristics are time-invariant, but there are some key message (or environmental) characteristics which may change over time. For example:
  -- source IP addresses may get added to DNSBLs
  -- message body URIs may get added to SURBLs
  -- the DCC/Razor/Pyzor status of a message may change
  -- users may change their spam threshold (or other prefs)
- Given that, it becomes pretty easy to envision a server-side software agent which might crawl unread user mail, "rescoring" or rechecking unread messages that earlier passed a scoring spam filter such as SpamAssassin.
- If a message initially passed SpamAssassin scoring, but then flunked a rescore, I believe the message's spam status can (and should) be updated appropriately.

Maybe We Should Filter *Just* Before Reading?

- In fact, arguably, the optimal time to do final filtering might be just before the user is about to read their mail. Why?...
  -- you've waited as long as you can for more reputation data
  -- you know that the user is actually trying to read their mail (some accounts may have mail that never gets read)
- On the other hand, we also know that:
  -- user access to mail often presents peak load issues (e.g., providers all have "prime time" peak load periods when seemingly "everyone" wants to access their email at the same time); you don't want to add load or introduce delays to what's already a busy time
  -- rescoring is the perfect sort of "garbage collection" maintenance activity to do during known low-load times
  -- you're going to needlessly wait to dump a lot of spam
- Is there a better time to rescore? Maybe event transitions?

Event Transitions

- Event transitions are state changes such as:
  -- a connecting IP address goes from unknown/presumed-to-be-clean status to known-to-be-a-spammer or known-to-be-a-spam-zombie status (e.g., the IP address gets listed on the Spamhaus SBL or XBL)
  -- a message body URI gets listed on the SURBL or URIBL
  -- message checksum get listed as bad on a collaborative filtering service, etc.
- As soon as an event transition of that sort occurs and that change of status becomes known to you, that's when you should update the message status of all related messages.
- Clearly, however, you can't sequentially rescan all messages every time an event transition occurs – event transitions occur virtually continuously (or at least every hour or few hours in the form of periodic zone file updates for DNSBLs)

Facilitating Event-Transition Filtering

- To make event-transition-triggered filtering realistically possible, we need to borrow an architectural lesson from Usenet News (NNTP).
- For those who may not be familiar with Usenet, INN (for example) creates an "overview" record with distilled information about each article. I'm proposing that every time we receive a mail message, providers should create a similar "mail overviews" database entry for that message.
- By distilling key information for each message into such a consolidated/indexed database, we eliminate the need to rescan individual message (unless we wanted to do so for some reason), thereby eliminating both the overhead associated with scanning large messages as well as the file system overhead associated with walking a large numbers of little teeny-tiny messages.

What Might Be In A Mail Overviews File?

- That "overviews" entry could contain some/all of:
  -- the message ID (or other "unique" identifier)
  -- a maildir path for the message (or equivalent storage ref)
  -- the mail "folder" associated with the message currently (inbox, spam folder, other automatically selected folder)
  -- the IP address of the system that connected and handed us that message (and the ASN of that IP address)
  -- the date/time the message was received
  -- the message envelop sender and message body sender
  -- any message body URIs (for SURBL filtering), as well as the IP addresses of those URIs (and the ASN of those IP addresses, their nameservers, registrar, etc.)
  -- any message body attachment checksums
  -- spam score components (for SpamAssassin, etc.)
  -- read-by-user flag, flagged-as-spam-by-user flag, etc.
  -- a journal of changes to the message's status over time

The Role of That Overviews File

- Once your server has an overview file in place, system access to mail messages (for web email, POP, IMAP, etc.) can occur via that mail overviews abstraction layer, but user email presentation would be unchanged. [Note: at least one IMAP server already creates indices similar to what I'm proposing, although not typically ones which contain directly spam-relevant information (e.g., see: http://www.dovecot.org/doc/mail-storages.txt)]
- So what if the system learns of a URI that's been listed as being spammy? That event could trigger updates to the spam status of all the messages associated with that URI, including potentially triggering other actions such as refoldering or /dev/null'ing the messages, etc.
- Email overviews also lay the foundation for easy data mining, exposing latent spam relationships which may be present.

So Why Isn't Everyone Doing Post-Delivery Filtering Right Now?

- Excellent question… Possible reasons include…
  -- some folks probably already are quietly doing this
  -- other folks haven't needed to do post-delivery filtering yet
  -- others might do it, if this technique was implemented in the mail software (or hardware appliance) which they use
  -- some providers have users who treat mail like an instant messaging application, downloading their new mail every minute, which could obviously limit potential applicability
  -- still others may worry about the so-called "disappearing message" problem
  -- there may also be concern about potential legal issues -- a message, once delivered to the customer's inbox, may be considered to be accessible solely by the customer
  -- "post-delivery filtering will cause a flood of bad bounces"
- Let's consider just a couple of those potential issues…

Users Who POP Their Mail "Every Minute"

- We all know that there some very "type A" personalities who POP their email every minute (including some of the folks in this room, I suspect!). Obviously, in that case, mail may not sit on the server very long, and there may not be much time to do any post-delivery filtering. So what about that, eh?
- Argument 1: Even for that sort of degenerate case, the user experience is no worse with post-delivery filtering than it would be w/o post-delivery filtering.
- Argument 2: You might be able to incent users to adopt less aggressive POP settings by showing the spam reduction that's possible for longer post-delivery filtering windows.
- Argument 3: Many users are not like "us" and use email casually, only occasionally connecting via a web email interface. Post-delivery filtering works great in that scenario.
- We need hard data on user access timing to really know…

(Potential) "Disappearing Message" Issue

- In talking with some colleagues about these ideas, the other issue that comes up, perhaps more than anything else, is the "disappearing message" problem. To briefly summarize:
  -- assume a user checks her inbox (but doesn't read any of their messages); she notices she has 36 messages
  -- assume post-delivery filtering works on that pending mail, and by the time the user logs back on to actually read her messages, there are only a dozen messages left (the others having been post-delivery filtered as spam)
  -- user gets worried… "Where did my other 24 pending messages go?" ==> user calls help desk ==> that costs $.
- Thus, either the interface must carefully explain what's going on so that there's zero chance for confusion (and support calls), or we need to move the *"messages are frozen, not eligible for further potential filtering"* frontier to encompass all messages present on the account as of the last user access.

Frozen-From-Further-Filtering-Frontier (F^5)

- That is, if you assume that it is hard to provide "push" status information in band ("We've identified and moved 24 messages from your inbox to your spam folder since you last logged in"), you really need to make sure you honor the frozen-from-further-filtering-frontier (F^5) to avoid presenting a user with potentially confusing information about the status of messages received on their account.
- If it turns out that you do need to honor F^5 (and most will):
  -- once a user logs in or otherwise accesses their account, all messages on their account as of that time, and any new messages that come in while they're logged in, become ineligible for any further post-delivery filtering
  -- once the user logs out, any new messages received after that point can get scrutinized and post-delivery filtere

  [conceptually, one could potentially imagine allowing users to explicitly "opt-in" to filtering at any time, even post F^5]

(Potential) Legal Issues

- I am not a lawyer and this talk isn't legal advice.
- Fortunately, you all DO have laywers, and this is a nice easily framed and interesting legal question: "If a user hasn't yet read mail we've delivered to their inbox, can we legally spam filter those still-unread messages, assuming we wanted to?"
- Having MAAWG meet in Brussels is also good in that it reminds us all that there are multiple legal systems governing ISP practices with respect to spam, and those systems are often not fully harmonized, so even if something were to be permissible (or forbidden) in one location, it would not necessarily also be permissible (or forbidden) in another.
- All that said, suitable terms of service and informed user consent can immunize providers against a tremendous number of potential legal issues. I'd urge you to be very candid with your customers about what you do.

(Potential) Bounce Problem

- When processing spam after the delivering host disconnects (as we would be doing under this proposal), another question that can come up is "What do we do with the spammy mail once we've identified it as unwanted AFTER the remote host has disconnected? It's too late then to just refuse/reject the mail outright…" So what could you do?

  -- Tag it as spam/move it to a spam folder? Absolutely.

  -- Silently delete it outright? We all know that some folks do this for at least some classes of unwanted mail (such as viral email) whether it is RFC compliant or not, but this is a riskier strategy in the case of "just" spam.

  -- What you absolutely MUST NOT DO is bounce an unwanted spam message to the "apparent sender" (obviously this could enable horrible DDoS possibilities)

Summary/Recommendations

- If you aren't doing post-delivery filtering, consider it.
- Part of your evaluation process should include a discussion with legal counsel about filtering-before-reading-but-after delivery-has-taken-place.
- If you elect to begin doing post-delivery filtering, construct overviews (index structures) for message traffic.
- You may be able to productively data mine that overview data for spam intelligence, assuming privacy policies permit.
- Update overviews and/or filter messages when event transitions occur.
- If possible (e.g., in a web email environment), make sure the user interface keeps the user informed about any spam "garbage collection" that has taken place, otherwise heed F^5
- Do NOT bounce any mail that you post-delivery filter.